

1/26

Polynucleotide and deduced amino acid sequence of hMLH1

```

-40          -20          1
.
GTTGAACATCTAGACGTTTCCTTGGCTCTTCTGGCGCCAAAATGTCGTTTCGTGGCAGGGG
+-----+-----+-----+-----+-----+-----+
CAACTTGTAGATCTGCAAAGGAACCGAGAAGACCGCGGTTTTACAGCAAGCACCGTCCCC
                                M S F V A G V
20          40          60
.
TTATTCGGCGGCTGGACGAGACAGTGGTGAACCGCATCGCGGCGGGGGAAGTTATCCAGC
+-----+-----+-----+-----+-----+-----+
AATAAGCCGCCGACCTGCTCTGTCACCACTTGGCGTAGCGCCGCCCCCTTCAATAGGTCG
  I R R L D E T V V N R I A A G E V I Q R
80          100         120
.
GGCCAGCTAATGCTATCAAAGAGATGATTGAGAACTGTTTAGATGCAAAATCCACAAGTA
+-----+-----+-----+-----+-----+-----+
CCGGTCGATTACGATAGTTTCTCTACTAACTCTTGACAAATCTACGTTTTAGGTGTTTCAT
  P A N A I K E M I E N C L D A K S T S I
140         160         180
.
TTCAAGTGATTGTTAAAGAGGGAGGCCTGAAGTTGATTTCAGATCCAAGACAATGGCACCG
+-----+-----+-----+-----+-----+-----+
AAGTTCACTAACAATTTCTCCCTCCGACTTCAACTAAGTCTAGGTTCTGTTACCGTGGC
  Q V I V K E G G L K L I Q I Q D N G T G
200         220         240
.
GGATCAGGAAAGAAGATCTGGATATTGTATGTGAAAGGTTCACTACTAGTAAACTGCAGT
+-----+-----+-----+-----+-----+-----+
CCTAGTCCTTTCTTCTAGACCTATAACATACACTTTCCAAGTGATGATCATTTGACGTCA
  I R K E D L D I V C E R F T T S K L Q S
260         280         300
.
CCTTTGAGGATTTAGCCAGTATTTCTACCTATGGCTTTTCGAGGTGAGGCTTTGGCCAGCA
+-----+-----+-----+-----+-----+-----+
GGAAACTCCTAAATCGGTCATAAAGATGGATACCGAAAGCTCCACTCCGAAACCGGTCGT
  F E D L A S I S T Y G F R G E A L A S I
320         340         360
.
TAAGCCATGTGGCTCATGTTACTATTACAACGAAAACAGCTGATGGAAAGTGTGCATACA
+-----+-----+-----+-----+-----+-----+
ATTCGGTACACCGAGTACAATGATAATGTTGCTTTTGTGCGACTACCTTTCACACGTATGT
  S H V A H V T I T T K T A D G K C A Y R

```

FIG. 1A

380 400 420
GAGCAAGTTACTCAGATGGAAAACCTGAAAGCCCCCTCCTAAACCATGTGCTGGCAATCAAG
+-----+-----+-----+-----+-----+-----+-----+-----+
CTCGTTCAATGAGTCTACCTTTTGACTTTCGGGGAGGATTTGGTACACGACCGTTAGTTC
A S Y S D G K L K A P P K P C A G N Q G
440 460 480
GGACCCAGATCACGGTGGAGGACCTTTTTTACAACATAGCCACGAGGAGAAAAGCTTTAA
+-----+-----+-----+-----+-----+-----+-----+-----+
CCTGGGTCTAGTGCCACCTCCTGGAAAAAATGTTGTATCGGTGCTCCTCTTTTCGAAATT
T Q I T V E D L F Y N I A T R R K A L K
500 520 540
AAAATCCAAGTGAAGAATATGGGAAAATTTTGAAGTTGTTGGCAGGTATTCAGTACACA
+-----+-----+-----+-----+-----+-----+-----+-----+
TTTAGGTTCACTTCTTATACCCTTTTAAACCTTCAACAACCGTCCATAAGTCATGTGT
N P S E E Y G K I L E V V G R Y S V H N
560 580 600
ATGCAGGCATTAGTTTCTCAGTTAAAAACAAGGAGAGACAGTAGCTGATGTTAGGACAC
+-----+-----+-----+-----+-----+-----+-----+-----+
TACGTCCGTAATCAAAGAGTCAATTTTTTGTTCCTCTCTGTCACTACAATCCTGTG
A G I S F S V K K Q G E T V A D V R T L
620 640 660
TACCCAATGCCTCAACCGTGGACAATATTCGCTCCGTCTTTGGAAATGCTGTTAGTCGAG
+-----+-----+-----+-----+-----+-----+-----+-----+
ATGGGTTACGGAGTTGGCACCTGTTATAAGCGAGGCAGAAACCTTTACGACAATCAGCTC
P N A S T V D N I R S V F G N A V S R E
680 700 720
AACTGATAGAAATTGGATGTGAGGATAAAACCCTAGCCTTCAAAATGAATGGTTACATAT
+-----+-----+-----+-----+-----+-----+-----+-----+
TTGACTATCTTTAACCTACACTCCTATTTTGGGATCGGAAGTTTACTTACCAATGTATA
L I E I G C E D K T L A F K M N G Y I S
740 760 780
CCAATGCAAATACTCAGTGAAGAAGTGCATCTTCTTACTCTTCATCAACCATCGTCTGG
+-----+-----+-----+-----+-----+-----+-----+-----+
GGTTACGTTTGTATGAGTCACTTCTTCACGTAGAAGAATGAGAAGTAGTTGGTAGCAGACC
N A N Y S V K K C I F L L F I N H R L V

FIG. 1B

800 820 840

 TAGAATCAACTTCCTTGAGAAAAGCCATAGAAACAGTGTATGCAGCCTATTTGCCCAAAA
 +-----+-----+-----+-----+-----+-----+
 ATCTTAGTTGAAGGAACTCTTTTCGGTATCTTTGTACATACGTCGGATAAACGGGTTTT
 E S T S L R K A I E T V Y A A Y L P K N
 860 880 900

 ACACACACCCATTCTGTACCTCAGTTTAGAAATCAGTCCCCAGAATGTGGATGTTAATG
 +-----+-----+-----+-----+-----+-----+
 TGTGTGTGGGTAAGGACATGGAGTCAAATCTTTAGTCAGGGGTCTTACACCTACAATTAC
 T H P F L Y L S L E I S P Q N V D V N V
 920 940 960

 TGCACCCACAAAGCATGAAGTTCACCTTCCTGCACGAGGAGAGCATCCTGGAGCGGGTGC
 +-----+-----+-----+-----+-----+-----+
 ACGTGGGGTGTTCGTACTTCAAGTGAAGGACGTGCTCCTCTCGTAGGACCTCGCCCACG
 H P T K H E V H F L H E E S I L E R V Q
 980 1000 1020

 AGCAGCACATCGAGAGCAAGCTCCTGGGCTCCAATTCCTCCAGGATGTACTTCACCCAGA
 +-----+-----+-----+-----+-----+-----+
 TCGTCGTGTAGCTCTCGTTCGAGGACCCGAGGTTAAGGAGGTCTTACATGAAGTGGGTCT
 Q H I E S K L L G S N S S R M Y F T Q T
 1040 1060 1080

 CTTTGCTACCAGGACTTGCTGGCCCCTCTGGGGAGATGGTTAAATCCACAACAAGTCTGA
 +-----+-----+-----+-----+-----+-----+
 GAAACGATGGTCCTGAACGACCGGGGAGACCCCTCTACCAATTTAGGTGTTGTTTCAGACT
 L L P G L A G P S G E M V K S T T S L T
 1100 1120 1140

 CCTCGTCTTCTACTTCTGGAAGTAGTGATAAGGTCTATGCCCACCAGATGGTTCGTACAG
 +-----+-----+-----+-----+-----+-----+
 GGAGCAGAAGATGAAGACCTTCATCACTATTCAGATACGGGTGGTCTACCAAGCATGTC
 S S S T S G S S D K V Y A H Q M V R T D
 1160 1180 1200

 ATCCCCGGAACAGAAGCTTGATGCATTTCTGCAGCCTCTGAGCAAACCCCTGTCCAGTC
 +-----+-----+-----+-----+-----+-----+
 TAAGGGCCCTTGCTTCGAACTACGTAAAGACGTGCGAGACTCGTTTGGGGACAGGTCAG
 S R E Q K L D A F L Q P L S K P L S S Q

FIG. 1C

1220 1240 1260

 AGCCCCAGGCCATTGTCACAGAGGATAAGACAGATATTTCTAGTGGCAGGGCTAGGCAGC
 +-----+-----+-----+-----+-----+-----+
 TCGGGGTCCGGTAACAGTGTCTCCTATTCTGTCTATAAAGATCACCGTCCCGATCCGTCG
 P Q A I V T E D K T D I S S G R A R Q Q
 1280 1300 1320

 AAGATGAGGAGATGCTTGAACCTCCAGCCCCTGCTGAAGTGGCTGCCAAAAATCAGAGCT
 +-----+-----+-----+-----+-----+-----+
 TTCTACTCCTCTACGAACTTGAGGGTCGGGGACGACTTCACCGACGGTTTTTAGTCTCGA
 D E E M L E L P A P A E V A A K N Q S L
 1340 1360 1380

 TGGAGGGGGATACAACAAAGGGGACTTCAGAAATGTCAGAGAAGAGAGGACCTACTTCCA
 +-----+-----+-----+-----+-----+-----+
 ACCTCCCCCTATGTTGTTTCCCCTGAAGTCTTTACAGTCTCTTCTCTCTGATGAAGGT
 E G D T T K G T S E M S E K R G P T S S
 1400 1420 1440

 GCAACCCAGAAAAGAGACATCGGGAAGATTCTGATGTGGAAATGGTGGAAAGATGATTCCC
 +-----+-----+-----+-----+-----+-----+
 CGTTGGGGTCTTTCTCTGTAGCCCTTCTAAGACTACACCTTTACCACCTTCTACTAAGGG
 N P R K R H R E D S D V E M V E D D S R
 1460 1480 1500

 GAAAGGAAATGACTGCAGCTTGTACCCCCCGGAGAAGGATCATTAACCTCACTAGTGT
 +-----+-----+-----+-----+-----+-----+
 CTTTCCTTTACTGACGTCGAACATGGGGGGCCTCTTCCTAGTAATTGGAGTGATCACAAA
 K E M T A A C T P R R R I I N L T S V L
 1520 1540 1560

 TGAGTCTCCAGGAAGAAATTAATGAGCAGGGACATGAGGTTCTCCGGGAGATGTTGCATA
 +-----+-----+-----+-----+-----+-----+
 ACTCAGAGGTCCTTCTTTAATTACTCGTCCCTGTACTCCAAGAGGCCCTCTACAACGTAT
 S L Q E E I N E Q G H E V L R E M L H N
 1580 1600 1620

 ACCACTCCTTCGTGGGCTGTGTGAATCCTCAGTGGGCCTTGGCACAGCATCAAACCAAGT
 +-----+-----+-----+-----+-----+-----+
 TGGTGAGGAAGCACCCGACACACTTAGGAGTCACCCGGAACCGTGTTCGTAGTTTGGTTCA
 H S F V G C V N P O W A L A O H O T K L

FIG. 1D

5/26

```

1640                      1660                      1680
.
TATACCTTCTCAACACCACCAAGCTTAGTGAAGAACTGTTCTACCAGATACTCATTTATG
+-----+-----+-----+-----+-----+-----+
ATATGGAAGAGTTGTGGTGGTTTCGAATCACTTCTTGACAAGATGGTCTATGAGTAAATAC
  Y  L  L  N  T  T  K  L  S  E  E  L  F  Y  Q  I  L  I  Y  D
1700                      1720                      1740
.
ATTTTGCCAATTTTGGTGTTCAGGTTATCGGAGCCAGCACCGCTCTTTGACCTTGCCA
+-----+-----+-----+-----+-----+-----+
TAAAACGGTTAAAACCACAAGAGTCCAATAGCCTCGGTCTGGCGAGAACTGGAACGGT
  F  A  N  F  G  V  L  R  L  S  E  P  A  P  L  F  D  L  A  M
1760                      1780                      1800
.
TGCTTGCCCTTAGATAGTCCAGAGAGTGGCTGGACAGAGGAAGATGGTCCCAAAGAAGGAC
+-----+-----+-----+-----+-----+-----+
ACGAACGGAATCTATCAGGTCTCTCACCAGCTGTCTCCTTCTACCAGGGTTTCTTCCTG
  L  A  L  D  S  P  E  S  G  W  T  E  E  D  G  P  K  E  G  L
1820                      1840                      1860
.
TTGCTGAATACATTGTTGAGTTTCTGAAGAAGAAGGCTGAGATGCTTGCAGACTATTTCT
+-----+-----+-----+-----+-----+-----+
AACGACTTATGTAACAACCTCAAAGACTTCTTCTTCCGACTCTACGAACGTCTGATAAAGA
  A  E  Y  I  V  E  F  L  K  K  K  A  E  M  L  A  D  Y  F  S
1880                      1900                      1920
.
CTTTGGAAATTGATGAGGAAGGGAACCTGATTGGATTACCCCTTCTGATTGACAACTATG
+-----+-----+-----+-----+-----+-----+
GAAACCTTTAACTACTCCTTCCCTTGGACTAACCTAATGGGGAAGACTAACTGTTGATAC
  L  E  I  D  E  E  G  N  L  I  G  L  P  L  L  I  D  N  Y  V
1940                      1960                      1980
.
TGCCCCCTTTGGAGGGACTGCCTATCTTCATTCTTCGACTAGCCACTGAGGTGAATTGGG
-----+-----+-----+-----+-----+-----+
ACGGGGGAAACCTCCCTGACGGATAGAAGTAAGAAGCTGATCGGTGACTCCACTTAACCC
  P  P  L  E  G  L  P  I  F  I  L  R  L  A  T  E  V  N  W  D
2000                      2020                      2040
.
ACGAAGAAAAGGAATGTTTTGAAAGCCTCAGTAAAGAATGCGCTATGTTCTATTCCATCC
+-----+-----+-----+-----+-----+-----+
TGCTTCTTTTCTTACAAAACCTTTCGGAGTCATTTCTTACGCGATACAAGATAAGGTAGG
  E  E  K  E  C  F  E  S  L  S  K  E  C  A  M  F  Y  S  I  R

```

FIG. 1E

6/26

```
2060                2080                2100
.                  .                  .
GGAAGCAGTACATATCTGAGGAGTCGACCCTCTCAGGCCAGCAGAGTGAAGTGCCTGGCT
+-----+-----+-----+-----+-----+-----+
CCTTCGTCATGTATAGACTCCTCAGCTGGGAGAGTCCGGTCGTCTCACTTCACGGACCGA
  K  Q  Y  I  S  E  E  S  T  L  S  G  Q  Q  S  E  V  P  G  S
2120                2140                2160
.                  .                  .
CCATTCCAAACTCCTGGAAGTGGACTGTGGAACACATTGTCTATAAAGCCTTGCGCTCAC
+-----+-----+-----+-----+-----+-----+
GGTAAGGTTTGAGGACCTTCACCTGACACCTTGTGTAACAGATATTTCGGAACGCGAGTG
  I  P  N  S  W  K  W  T  V  E  H  I  V  Y  K  A  L  R  S  H
2180                2200                2220
.                  .                  .
ACATTCTGCCTCCTAAACATTTACAGAAGATGGAAATATCCTGCAGCTTGCTAACCTGC
+-----+-----+-----+-----+-----+-----+
TGTAAGACGGAGGATTTGTAAAGTGTCTTCTACCTTTATAGGACGTCGAACGATTGGACG
  I  L  P  P  K  H  F  T  E  D  G  N  I  L  Q  L  A  N  L  P
2240                2260                2280
.                  .                  .
CTGATCTATACAAAGTCTTTGAGAGGTGTTAAATATGGTTATTTATGCACTGTGGGATGT
+-----+-----+-----+-----+-----+-----+
GACTAGATATGTTTCAGAACTCTCCACAATTTATACCAATAAATACGTGACACCCTACA
  D  L  Y  K  V  F  E  R  C  *
2300                2320                2340
.                  .                  .
GTTCTTCTTTCTCTGTATTCCGATACAAAGTGTTGTATCAAAGTGTGATATACAAAGTGT
+-----+-----+-----+-----+-----+-----+
CAAGAAGAAAGAGACATAAGGCTATGTTTCACAACATAGTTTCACACTATATGTTTCACA
2360                2380                2400
.                  .                  .
ACCAACATAAGTGTTGGTAGCACTTAAGACTTATACTTGCCTTCTGATAGTATTCCTTTA
+-----+-----+-----+-----+-----+-----+
TGGTTGTATTTCACAACCATCGTGAATTCTGAATATGAACGGAAGACTATCATAAGGAAAT
2420                2440                2460
.                  .                  .
TACACAGTGGATTGATTATAAATAAATAGATGTGTCTTAACATAAAAAAAAAAAAAAAAAA
+-----+-----+-----+-----+-----+-----+
ATGTGTCACCTAACTAATATTTATTTATCTACACAGAATTGTATTTTTTTTTTTTTTTTTT

2480
.
AAAAA
+----
TTTTT
```

FIG. 1F

7/26

Polynucleotide and deduced amino acid sequence of hMLH2

```

-70          -50          -30
.           .           .
GGCACGAGTGGCTGCTTGC GGCTAGTGGATGGTAATTGCCTGCCTCGCGCTAGCAGCAAG
-----+-----+-----+-----+-----+-----+-----+
CCGTGCTCACCGACGAACGCCGATCACCTACCATTAACGGACGGAGCGCGATCGTCGTTC
-10          10          30
.           .           .
CTGCTCTGT TAAAAGCGAAAATGAAACAATTGCCTGCGGCAACAGTTCGACTCCTTTCAA
-----+-----+-----+-----+-----+-----+
GACGAGACAATTTTCGCTTTTACTTTGT TAAACGGACGCCGTTGTCAAGCTGAGGAAAGTT
                    M K Q L P A A T V R L L S S
                50          70          90
.           .           .
GTTCTCAGATCATCACTTCGGTGGTCAGTGT TGTAAAAGAGCTTATTGAAAAC T CCTTGG
-----+-----+-----+-----+-----+-----+
CAAGAGTCTAGTAGTGAAGCCACCAGTCACAACATTTTCTCGAATAACTTTTGAGGAACC
S Q I I T S V V S V V K E L I E N S L D
                110          130          150
.           .           .
ATGCTGGTGCCACAAGCGTAGATGTTAAACTGGAGAACTATGGATTTGATAAAATTGAGG
-----+-----+-----+-----+-----+-----+
TACGACCACGGTGTTCGCATCTACAATTTGACCTCTTGATACCTAAACTATTTTAACTCC
A G A T S V D V K L E N Y G F D K I E V
                170          190          210
.           .           .
TGCGAGATAACGGGGAGGGTATCAAGGCTGTTGATGCACCTGTAATGGCAATGAAGTACT
-----+-----+-----+-----+-----+-----+
ACGCTCTATTGCCCCCTCCCATAGTTCCGACAAC TACGTGGACATTACCGTTACTTCATGA
R D N G E G I K A V D A P V M A M K Y Y
                230          250          270
.           .           .
ACACCTCAAAAATAAATAGTCATGAAGATCTTGAAAATTTGACAAC TACGGTTTTTCGTG
-----+-----+-----+-----+-----+-----+
TGTGGAGTTTTTTATTTATCAGTACTTCTAGAACTTTTAAACTGTTGAATGCCAAAAGCAC
T S K I N S H E D L E N L T T Y G F R G
                290          310          330
.           .           .
GAGAAGCCTTGGGGTCAATTTGTTGTATAGCTGAGGTTTTAATTACAACAAGAACGGCTG
-----+-----+-----+-----+-----+-----+
CTCTTCGGAACCCCAGTTAAACAACATATCGACTCCAAAATTAATGTTGTTCTTGCCGAC
E A L G S I C C I A E V L I T T R T A A

```

FIG. 2A

8/26

```

      350              370              390
      .               .               .
CTGATAATTTTAGCACCCAGTATGTTTTAGATGGCAGTGGCCACATACTTTCTCAGAAAC
-----+-----+-----+-----+-----+-----+-----+
GACTATTAAAATCGTGGGTCATACAAAATCTACCGTCACCGGTGTATGAAAGAGTCTTTG
  D  N  F  S  T  Q  Y  V  L  D  G  S  G  H  I  L  S  Q  K  P
    410              430              450

      .               .               .
CTTCACATCTTGGTCAAGGTACAACGTGTAAGTCTTTAAGATTATTTAAGAATCTACCTG
-----+-----+-----+-----+-----+-----+-----+
GAAGTGTAGAACCAGTTCCATGTTGACATTGACGAAATTCTAATAAATTCTTAGATGGAC
  S  H  L  G  Q  G  T  T  V  T  A  L  R  L  F  K  N  L  P  V
    470              490              510

      .               .               .
TAAGAAAGCAGTTTTACTCAACTGCAAAAAAATGTAAAGATGAAATAAAAAAGATCCAAG
-----+-----+-----+-----+-----+-----+-----+
ATTCTTTTCGTCAAAATGAGTTGACGTTTTTTTACATTTCTACTTTATTTTTTCTAGGTTT
  R  K  Q  F  Y  S  T  A  K  K  C  K  D  E  I  K  K  I  Q  D
    530              550              570

      .               .               .
ATCTCCTCATGAGCTTTGGTATCCTTAAACCTGACTTAAGGATTGTCTTTGTACATAACA
-----+-----+-----+-----+-----+-----+-----+
TAGAGGAGTACTCGAAACCATAGGAATTTGGACTGAATTCCTAACAGAAACATGTATTGT
  L  L  M  S  F  G  I  L  K  P  D  L  R  I  V  F  V  H  N  K
    590              610              630

      .               .               .
AGGCAGTTATTTGGCAGAAAAGCAGAGTATCAGATCACAAGATGGCTCTCATGTCAGTTC
-----+-----+-----+-----+-----+-----+-----+
TCCGTCAATAAACCGTCTTTTCGTCTCATAGTCTAGTGTTCTACCGAGAGTACAGTCAAG
  A  V  I  W  Q  K  S  R  V  S  D  H  K  M  A  L  M  S  V  L
    650              670              690

      .               .               .
TGGGGACTGCTGTTATGAACAATATGGAATCCTTTCAGTACCACTCTGAAGAATCTCAGA
-----+-----+-----+-----+-----+-----+-----+
ACCCCTGACGACAATACTTGTTATACCTTAGGAAAGTCATGGTGAGACTTCTTAGAGTCT
  G  T  A  V  M  N  N  M  E  S  F  Q  Y  H  S  E  E  S  Q  I
    710              730              750

      .               .               .
TTTATCTCAGTGGATTTCTTCCAAAGTGTGATGCAGACCACTCTTTCAGTCTTTCAA
-----+-----+-----+-----+-----+-----+-----+
AAATAGAGTCACCTAAAGAAGGTTTTCACACTACGTCTGGTGAGAAAGTGATCAGAAAGTT
  Y  L  S  G  F  L  P  K  C  D  A  D  H  S  F  T  S  L  S  T

```

FIG. 2B

9/26

```

      770              790              810
      .               .               .
CACCAGAAAGAAGTTTCATCTTCATAAACAGTCGACCAGTACATCAAAAAGATATCTTAA
-----+-----+-----+-----+-----+-----+-----+
GTGGTCTTTCTTCAAAGTAGAAGTATTTGTCAGCTGGTCATGTAGTTTTTCTATAGAATT
  P  E  R  S  F  I  F  I  N  S  R  P  V  H  Q  K  D  I  L  K
      830              850              870

      .               .               .
AGTTAATCCGACATCATTACAATCTGAAATGCCTAAAGGAATCTACTCGTTTGTATCCTG
-----+-----+-----+-----+-----+-----+-----+
TCAATTAGGCTGTAGTAATGTTAGACTTTACGGATTTTCCTTAGATGAGCAAACATAGGAC
  L  I  R  H  H  Y  N  L  K  C  L  K  E  S  T  R  L  Y  P  V
      890              910              930

      .               .               .
TTTTCTTTCTGAAAATCGATGTTCTACAGCTGATGTTGATGTAAATTTAACACCAGATA
-----+-----+-----+-----+-----+-----+-----+
AAAAGAAAGACTTTTAGCTACAAGGATGTCGACTACAACACTACATTTAAATTGTGGTCTAT
  F  F  L  K  I  D  V  P  T  A  D  V  D  V  N  L  T  P  D  K
      950              970              990

      .               .               .
AAAGCCAAGTATTATTACAAAATAAGGAATCTGTTTTAATTGCTCTTGAAAATCTGATGA
-----+-----+-----+-----+-----+-----+-----+
TTTCGGTTCATAATAATGTTTTATTCCCTTAGACAAAATTAACGAGAACTTTTAGACTACT
  S  Q  V  L  L  Q  N  K  E  S  V  L  I  A  L  E  N  L  M  T
      1010             1030             1050

      .               .               .
CGACTTGTTATGGACCATTACCTAGTACAAATTCTTATGAAAATAATAAACAGATGTTT
-----+-----+-----+-----+-----+-----+-----+
GCTGAACAATACCTGGTAATGGATCATGTTTAAGAATACTTTTATTATTTTGTCTACAAA
  T  C  Y  G  P  L  P  S  T  N  S  Y  E  N  N  K  T  D  V  S
      1070             1090             1110

      .               .               .
CCGCAGCTGACATCGTTCTTAGTAAAACAGCAGAAACAGATGTGCTTTTTTAATAAAGTGG
-----+-----+-----+-----+-----+-----+-----+
GGCGTCGACTGTAGCAAGAATCATTTTGTGCTCTTTGTCTACACGAAAAATTATTTACC
  A  A  D  I  V  L  S  K  T  A  E  T  D  V  L  F  N  K  V  E
      1130             1150             1170

      .               .               .
AATCATCTGGAAAGAATTATTCAAATGTTGATACTTCAGTCATTCCATTCCAAAATGATA
-----+-----+-----+-----+-----+-----+-----+
TTAGTAGACCTTTCTTAATAAGTTTACAACATATGAAGTCAGTAAGGTAAGGTTTACTAT
  S  S  G  K  N  Y  S  N  V  D  T  S  V  I  P  F  Q  N  D  M

```

FIG. 2C

1. *Chlorophyll a* (Chl *a*)
 2. *Chlorophyll b* (Chl *b*)
 3. *Chlorophyll c* (Chl *c*)
 4. *Chlorophyll d* (Chl *d*)
 5. *Chlorophyll e* (Chl *e*)
 6. *Chlorophyll f* (Chl *f*)
 7. *Chlorophyll g* (Chl *g*)
 8. *Chlorophyll h* (Chl *h*)
 9. *Chlorophyll i* (Chl *i*)
 10. *Chlorophyll j* (Chl *j*)
 11. *Chlorophyll k* (Chl *k*)
 12. *Chlorophyll l* (Chl *l*)
 13. *Chlorophyll m* (Chl *m*)
 14. *Chlorophyll n* (Chl *n*)
 15. *Chlorophyll o* (Chl *o*)
 16. *Chlorophyll p* (Chl *p*)
 17. *Chlorophyll q* (Chl *q*)
 18. *Chlorophyll r* (Chl *r*)
 19. *Chlorophyll s* (Chl *s*)
 20. *Chlorophyll t* (Chl *t*)
 21. *Chlorophyll u* (Chl *u*)
 22. *Chlorophyll v* (Chl *v*)
 23. *Chlorophyll w* (Chl *w*)
 24. *Chlorophyll x* (Chl *x*)
 25. *Chlorophyll y* (Chl *y*)
 26. *Chlorophyll z* (Chl *z*)
 27. *Chlorophyll aa* (Chl *aa*)
 28. *Chlorophyll ab* (Chl *ab*)
 29. *Chlorophyll ac* (Chl *ac*)
 30. *Chlorophyll ad* (Chl *ad*)
 31. *Chlorophyll ae* (Chl *ae*)
 32. *Chlorophyll af* (Chl *af*)
 33. *Chlorophyll ag* (Chl *ag*)
 34. *Chlorophyll ah* (Chl *ah*)
 35. *Chlorophyll ai* (Chl *ai*)
 36. *Chlorophyll aj* (Chl *aj*)
 37. *Chlorophyll ak* (Chl *ak*)
 38. *Chlorophyll al* (Chl *al*)
 39. *Chlorophyll am* (Chl *am*)
 40. *Chlorophyll an* (Chl *an*)
 41. *Chlorophyll ao* (Chl *ao*)
 42. *Chlorophyll ap* (Chl *ap*)
 43. *Chlorophyll aq* (Chl *aq*)
 44. *Chlorophyll ar* (Chl *ar*)
 45. *Chlorophyll as* (Chl *as*)
 46. *Chlorophyll at* (Chl *at*)
 47. *Chlorophyll au* (Chl *au*)
 48. *Chlorophyll av* (Chl *av*)
 49. *Chlorophyll aw* (Chl *aw*)
 50. *Chlorophyll ax* (Chl *ax*)
 51. *Chlorophyll ay* (Chl *ay*)
 52. *Chlorophyll az* (Chl *az*)
 53. *Chlorophyll aza* (Chl *aza*)
 54. *Chlorophyll abz* (Chl *abz*)
 55. *Chlorophyll acz* (Chl *acz*)
 56. *Chlorophyll adz* (Chl *adz*)
 57. *Chlorophyll aez* (Chl *aez*)
 58. *Chlorophyll afz* (Chl *afz*)
 59. *Chlorophyll agz* (Chl *agz*)
 60. *Chlorophyll ahz* (Chl *ahz*)
 61. *Chlorophyll aiz* (Chl *aiz*)
 62. *Chlorophyll ajz* (Chl *ajz*)
 63. *Chlorophyll akz* (Chl *akz*)
 64. *Chlorophyll alz* (Chl *alz*)
 65. *Chlorophyll amz* (Chl *amz*)
 66. *Chlorophyll anz* (Chl *anz*)
 67. *Chlorophyll aoz* (Chl *aoz*)
 68. *Chlorophyll apz* (Chl *apz*)
 69. *Chlorophyll aqz* (Chl *aqz*)
 70. *Chlorophyll arz* (Chl *arz*)
 71. *Chlorophyll asz* (Chl *asz*)
 72. *Chlorophyll atz* (Chl *atz*)
 73. *Chlorophyll auz* (Chl *auz*)
 74. *Chlorophyll avz* (Chl *avz*)
 75. *Chlorophyll awz* (Chl *awz*)
 76. *Chlorophyll axz* (Chl *axz*)
 77. *Chlorophyll ayz* (Chl *ayz*)
 78. *Chlorophyll ayz* (Chl *ayz*)
 79. *Chlorophyll azz* (Chl *azz*)
 80. *Chlorophyll azaa* (Chl *aza*)
 81. *Chlorophyll abz* (Chl *abz*)
 82. *Chlorophyll acz* (Chl *acz*)
 83. *Chlorophyll adz* (Chl *adz*)
 84. *Chlorophyll aez* (Chl *aez*)
 85. *Chlorophyll afz* (Chl *afz*)
 86. *Chlorophyll agz* (Chl *agz*)
 87. *Chlorophyll ahz* (Chl *ahz*)
 88. *Chlorophyll aiz* (Chl *aiz*)
 89. *Chlorophyll ajz* (Chl *ajz*)
 90. *Chlorophyll akz* (Chl *akz*)
 91. *Chlorophyll alz* (Chl *alz*)
 92. *Chlorophyll amz* (Chl *amz*)
 93. *Chlorophyll anz* (Chl *anz*)
 94. *Chlorophyll aoz* (Chl *aoz*)
 95. *Chlorophyll apz* (Chl *apz*)
 96. *Chlorophyll aqz* (Chl *aqz*)
 97. *Chlorophyll arz* (Chl *arz*)
 98. *Chlorophyll asz* (Chl *asz*)
 99. *Chlorophyll atz* (Chl *atz*)
 100. *Chlorophyll auz* (Chl *auz*)
 101. *Chlorophyll avz* (Chl *avz*)
 102. *Chlorophyll awz* (Chl *awz*)
 103. *Chlorophyll axz* (Chl *axz*)
 104. *Chlorophyll ayz* (Chl *ayz*)
 105. *Chlorophyll ayz* (Chl *ayz*)
 106. *Chlorophyll azz* (Chl *azz*)
 107. *Chlorophyll azaa* (Chl *aza*)
 108. *Chlorophyll abz* (Chl *abz*)
 109. *Chlorophyll acz* (Chl *acz*)
 110. *Chlorophyll adz* (Chl *adz*)
 111. *Chlorophyll aez* (Chl *aez*)
 112. *Chlorophyll afz* (Chl *afz*)
 113. *Chlorophyll agz* (Chl *agz*)
 114. *Chlorophyll ahz* (Chl *ahz*)
 115. *Chlorophyll aiz* (Chl *aiz*)
 116. *Chlorophyll ajz* (Chl *ajz*)
 117. *Chlorophyll akz* (Chl *akz*)
 118. *Chlorophyll alz* (Chl *alz*)
 119. *Chlorophyll amz* (Chl *amz*)
 120. *Chlorophyll anz* (Chl *anz*)
 121. *Chlorophyll aoz* (Chl *aoz*)
 122. *Chlorophyll apz* (Chl *apz*)
 123. *Chlorophyll aqz* (Chl *aqz*)
 124. *Chlorophyll arz* (Chl *arz*)
 125. *Chlorophyll asz* (Chl *asz*)
 126. *Chlorophyll atz* (Chl *atz*)
 127. *Chlorophyll auz* (Chl *auz*)
 128. *Chlorophyll avz* (Chl *avz*)
 129. *Chlorophyll awz* (Chl *awz*)
 130. *Chlorophyll axz* (Chl *axz*)
 131. *Chlorophyll ayz* (Chl *ayz*)
 132. *Chlorophyll ayz* (Chl *ayz*)
 133.

FIG. 2D

11/26

1610 1630 1650
CAATCCCTGAACAAATGAATCTTAATGAAGATTCATGTAACAAAAAATCAAATGTAATAG
-----+-----+-----+-----+-----+-----+-----+
GTTAGGGACTTGTTTACTTAGAATTACTTCTAAGTACATTGTTTTTTAGTTTACATTATC
I P E Q M N L N E D S C N K K S N V I D
1670 1690 1710
ATAATAAATCTGGAAAAGTTACAGCTTATGATTTACTTAGCAATCGAGTAATCAAGAAAC
-----+-----+-----+-----+-----+-----+-----+
TATTATTTAGACCTTTTCAATGTCGAATACTAAATGAATCGTTAGCTCATTAGTTCTTTG
N K S G K V T A Y D L L S N R V I K K P
1730 1750 1770
CCATGTCAGCAAGTGCTCTTTTTGTTCAAGATCATCGTCCTCAGTTTCTCATAGAAAATC
-----+-----+-----+-----+-----+-----+-----+
GGTACAGTCGTTACGAGAAAAACAAGTTCTAGTAGCAGGAGTCAAAGAGTATCTTTTAG
M S A S A L F V Q D H R P Q F L I E N P
1790 1810 1830
CTAAGACTAGTTTtagaggatgCAACACTACAAATTGAAGAACTGTGGAAGACATTGAGTG
-----+-----+-----+-----+-----+-----+-----+
GATTCTGATCAAATCTCCTACGTTGTGATGTTTAACTTCTTGACACCTTCTGTAACCTCAC
K T S L E D A T L Q I E E L W K T L S E
1850 1870 1890
AAGAGGAAAAACTGAAATATGAAGAGAAGGCTACTAAAGACTTGAACGATACAATAGTC
-----+-----+-----+-----+-----+-----+-----+
TTCTCCTTTTTGACTTTATACTTCTCTCCGATGATTTCTGAACCTTGCTATGTTATCAG
E E K L K Y E E K A T K D L E R Y N S Q
1910 1930 1950
AAATGAAGAGAGCCATTGAACAGGAGTCACAAATGTCACTAAAAGATGGCAGAAAAAAGA
-----+-----+-----+-----+-----+-----+-----+
TTTACTTCTCTCGGTAACCTGTCCTCAGTGTTTACAGTGATTTTCTACCGTCTTTTTTCT
M K R A I E Q E S Q M S L K D G R K K I
1970 1990 2010
TAAAACCCACCAGCGCATGGAATTTGGCCCAGAAGCACAAAGTTAAAACCTCATTATCTA
-----+-----+-----+-----+-----+-----+-----+
ATTTTGGGTGGTCGCGTACCTTAAACCGGTCTTCGTGTTCAATTTTTGGAGTAATAGAT
K P T S A W N L A Q K H K L K T S L S N

FIG. 2E

Figure 1 displays 12 histograms showing the distribution of the number of non-zero elements in the vector x for different values of n . The histograms are arranged in a 6x2 grid. The left column shows distributions for $n = 10, 20, 30, 40, 50, 60$. The right column shows distributions for $n = 70, 80, 90, 100, 110, 120$. Each histogram has 'Number of non-zero elements' on the x-axis (ranging from 0 to 120) and 'Frequency' on the y-axis (ranging from 0 to 100). The distributions are roughly bell-shaped and centered around 60-70 non-zero elements.

FIG. 2F

13/26

2450 2470 2490
CAGCGAATGGTTTCAAGATAAAATTGATACCAGGAGTTTCAATTACTGAAAATTACTTGG
-----+-----+-----+-----+-----+-----+-----+
GTCGCTTACCAAAGTTCTATTTTAACTATGGTCCTCAAAGTTAATGACTTTTAAATGAACC
A N G F K I K L I P G V S I T E N Y L E
2510 2530 2550
AAATAGAAGGAATGGCTAATTGTCTCCCATTTCTATGGAGTAGCAGATTTAAAAGAAATTC
-----+-----+-----+-----+-----+-----+-----+
TTTATCTTCCTTACCGATTAACAGAGGGTAAGATACCTCATCGTCTAAATTTTCTTTAAG
I E G M A N C L P F Y G V A D L K E I L
2570 2590 2610
TTAATGCTATATTAAACAGAAATGCAAAGGAAGTTTATGAATGTAGACCTCGCAAAGTGA
-----+-----+-----+-----+-----+-----+-----+
AATTACGATATAATTTGTCTTTACGTTTCCTTCAAATACTTACATCTGGAGCGTTTCACT
N A I L N R N A K E V Y E C R P R K V I
2630 2650 2670
TAAGTTATTTAGAGGGAGAAGCAGTGCGTCTATCCAGACAATTACCCATGTACTTATCAA
-----+-----+-----+-----+-----+-----+-----+
ATTCAATAAATCTCCCTCTTCGTCACGCAGATAGGTCTGTTAATGGGTACATGAATAGTT
S Y L E G E A V R L S R Q L P M Y L S K
2690 2710 2730
AAGAGGACATCCAAGACATTATCTACAGAATGAAGCACCAGTTTGGAAATGAAATTAAAG
-----+-----+-----+-----+-----+-----+-----+
TTCTCCTGTAGGTTCTGTAATAGATGTCTTACTTCGTGGTCAAACCTTTACTTTAATTTCT
E D I Q D I I Y R M K H Q F G N E I K E
2750 2770 2790
AGTGTGTTTCATGGTCGCCCATTTTTTCATCATTTAACCTATCTTCCAGAACTACATGAT
-----+-----+-----+-----+-----+-----+-----+
TCACACAAGTACCAGCGGGTAAAAAAGTAGTAAATTGGATAGAAGGTCTTTGATGTACTA
C V H G R P F F H H L T Y L P E T T *
2810 2830 2850
TAAATATGTTTAAAGAAGATTAGTTACCATTGAAATTGGTCTGTCATAAAACAGCATGAG
-----+-----+-----+-----+-----+-----+-----+
ATTTATACAAATTCCTCTAATCAATGGTAACTTTAACCAAGACAGTATTTTGTCTGCTACTC

FIG. 2G

[illegible]

2910

-----+-----+-----+-----+-----+

2970

GACTGAACAAAAATATAACTTTTTTCAAGGTGCATAACATCTTTTGCATTTATTTGATTA

— — —

TTG

FIG. 2H

15/26

Polynucleotide and deduced amino acid sequence of hMLH3

```

-20          0          20
.           .           .
CGAGGCGGATCGGGTGTTCATCCATGGAGCGAGCTGAGAGCTCGAGTACAGAACCTGCT
---+-----+-----+-----+-----+-----+-----+-----
GCTCCGCCTAGCCCACAACGTAGGTACCTCGCTCGACTCTCGAGCTCATGTCTTGGACGA
      M E R A E S S S T E P A
40          60          80
.           .           .
AAGGCCATCAAACCTATTGATCGGAAGTCAGTCCATCAGATTTGCTCTGGGCAGGTGGTA
---+-----+-----+-----+-----+-----+-----+-----
TTCCGGTAGTTTGGATAACTAGCCTTCAGTCAGGTAGTCTAAACGAGACCCGTCCACCAT
K A I K P I D R K S V H Q I C S G Q V V
100        120        140
.           .           .
CTGAGTCTAAGCACTGCGGTAAAGGAGTTAGTAGAAAACAGTCTGGATGCTGGTGCCACT
---+-----+-----+-----+-----+-----+-----+-----
GACTCAGATTTCGTGACGCCATTTCTCAATCATCTTTTGTGTCAGACCTACGACCACGGTGA
L S L S T A V K E L V E N S L D A G A T
160        180        200
.           .           .
AATATTGATCTAAAGCTTAAGGACTATGGAGTGGATCTTATTGAAGTTTCAGACAATGGA
---+-----+-----+-----+-----+-----+-----+-----
TTATAACTAGATTTTGAATTCCTGATACCTCACCTAGAATAACTTCAAAGTCTGTTACCT
N I D L K L K D Y G V D L I E V S D N G
220        240        260
.           .           .
TGTGGGGTAGAAGAAGAAAACCTCGAAGGCTTAACTCTGAAACATCACACATCTAAGATT
---+-----+-----+-----+-----+-----+-----+-----
ACACCCCATCTTCTTCTTTTGAAGCTTCCGAATTGAGACTTTGTAGTGTGTAGATTCTAA
C G V E E E N F E G L T L K H H T S K I
280        300        320
.           .           .
CAAGAGTTTGCCGACCTAACTCAGGTTGAAACTTTTGGCTTTCGGGGGAAGCTCTGAGC
---+-----+-----+-----+-----+-----+-----+-----
GTTCTCAAACGGGCTGGATTGAGTCCAACCTTTGAAAACCGAAAGCCCCCTTCGAGACTCG
Q E F A D L T Q V E T F G F R G E A L S
340        360        380
.           .           .
TCACTTTGTGCACTGAGCGATGTCACCATTTCTACCTGCCACGCATCGGCGAAGGTTGGA
---+-----+-----+-----+-----+-----+-----+-----
AGTGAAACACGTGACTCGCTACAGTGGTAAAGATGGACGGTGCGTAGCCGCTTCCAACCT
S L C A L S D V T I S T C H A S A K V G

```

FIG. 3A

16/26

400 420 440

ACTCGACTGATGTTTGATCACAATGGGAAAATTATCCAGAAAACCCCCTACCCCCGCCCC
 ---+-----+-----+-----+-----+-----+-----+-----
 TGAGCTGACTACAACTAGTGTTACCCTTTTAATAGGTCTTTTGGGGGATGGGGGCGGGG
 T R L M F D H N G K I I Q K T P Y P R P
 460 480 500

AGAGGGACCACAGTCAGCGTGCAGCAGTTATTTTCCACACTACCTGTGCGCCATAAGGAA
 ---+-----+-----+-----+-----+-----+-----+-----
 TCTCCCTGGTGTGTCAGTCGCACGTCGTCAATAAAAGGTGTGATGGACACGCGGTATTCCTT
 R G T T V S V Q Q L F S T L P V R H K E
 520 540 560

TTTCAAAGGAATATTAAGAAGGAGTATGCCAAAATGGTCCAGGTCTTACATGCATACTGT
 ---+-----+-----+-----+-----+-----+-----+-----
 AAAGTTTCCTTATAATTCTTCCTCATACGGTTTTACCAGGTCCAGAATGTACGTATGACA
 F Q R N I K K E Y A K M V Q V L H A Y C
 580 600 620

ATCATTTCAGCAGGCATCCGTGTAAGTTGCACCAATCAGCTTGGACAAGGAAAACGACAG
 ---+-----+-----+-----+-----+-----+-----+-----
 TAGTAAAGTCGTCCGTAGGCACATTCAACGTGGTGTAGTCGAACCTGTTTCCTTTTGCTGTC
 I I S A G I R V S C T N Q L G Q G K R Q
 640 660 680

CCTGTGGTATGCACAGGTGGAAGCCCCAGCATAAAGGAAAATATCGGCTCTGTGTTTGGG
 ---+-----+-----+-----+-----+-----+-----+-----
 GGACACCATACGTGTCCACCTTCGGGGTTCGTATTTTCCTTTTATAGCCGAGACACAAACCC
 P V V C T G G S P S I K E N I G S V F G
 700 720 740

CAGAAGCAGTTGCAAAGCCTCATTCCTTTTGTTCAGCTGCCCCCTAGTGACTCCGTGTGT
 ---+-----+-----+-----+-----+-----+-----+-----
 GTCTTCGTCAACGTTTCGGAGTAAGGAAAACAAGTCGACGGGGGATCACTGAGGCACACA
 Q K Q L Q S L I P F V Q L P P S D S V C
 760 780 800

GAAGAGTACGGTTTGAGCTGTTTCGGATGCTCTGCATAATCTTTTTTACATCTCAGGTTTC
 ---+-----+-----+-----+-----+-----+-----+-----
 CTTCTCATGCCAAACTCGACAAGCCTACGAGACGTATTAGAAAAAATGTAGAGTCCAAAG
 E E Y G L S C S D A L H N L F Y I S G F

FIG. 3B

17/26

820 840 860

ATTTCACAATGCACGCATGGAGTTGGAAGGAGTTCAACAGACAGACAGTTTTTCTTTATC
 ---+-----+-----+-----+-----+-----+-----+-----
 TAAAGTGTTACGTGCGTACCTCAACCTTCCTCAAGTTGTCTGTCTGTCAAAAAGAAATAG
 I S Q C T H G V G R S S T D R Q F F F I
 880 900 920

AACCGGCGGCCCTTGTGACCCAGCAAAGGTCTGCAGACTCGTGAATGAGGTCTACCACATG
 ---+-----+-----+-----+-----+-----+-----+-----
 TTGGCCGCGCGGAACACTGGGTGCTTTCCAGACGTCTGAGCACTTACTCCAGATGGTGTAC
 N R R P C D P A K V C R L V N E V Y H M
 940 960 980

TATAATCGACACCAGTATCCATTTGTTGTTCTTAACATTTCTGTTGATTCAGAATGCGTT
 ---+-----+-----+-----+-----+-----+-----+-----
 ATATTAGCTGTGGTCATAGGTAAACAACAAGAATTGTAAAGACAACCTAAGTCTTACGCAA
 Y N R H Q Y P F V V L N I S V D S E C V
 1000 1020 1040

GATATCAATGTTACTCCAGATAAAAGGCAAATTTTGCTACAAGAGGAAAAGCTTTTGTG
 ---+-----+-----+-----+-----+-----+-----+-----
 CTATAGTTACAATGAGGTCTATTTTCCGTTTAAAACGATGTTCTCCTTTTCGAAAACAAC
 D I N V T P D K R Q I L L Q E E K L L L
 1060 1080 1100

GCAGTTTTAAAGACCTCTTTGATAGGAATGTTTGATAGTGATGTCAACAAGCTAAATGTC
 ---+-----+-----+-----+-----+-----+-----+-----
 CGTCAAAATTTCTGGAGAACTATCCTTACAACTATCACTACAGTTGTTTCGATTTACAG
 A V L K T S L I G M F D S D V N K L N V
 1120 1140 1160

AGTCAGCAGCCACTGCTGGATGTTGAAGGTAACCTTAATAAAAATGCATGCAGCGGATTTG
 ---+-----+-----+-----+-----+-----+-----+-----
 TCAGTCGTCGGTGACGACCTACAACCTCCATTGAATTATTTTTACGTACGTGCGCTAAAC
 S Q Q P L L D V E G N L I K M H A A D L
 1180 1200 1220

GAAAAGCCCATGGTAGAAAAGCAGGATCAATCCCCTTCATTAAGGACTGGAGAAGAAAAA
 ---+-----+-----+-----+-----+-----+-----+-----
 CTTTTCGGGTACCATCTTTTCGTCCTAGTTAGGGGAAGTAATTCCTGACCTCTTCTTTTT
 E K P M V E K Q D Q S P S L R T G E E K

FIG. 3C

18/26

1240 1260 1280
AAAGACGTGTCCATTTCCAGACTGCGAGAGGCCTTTTCTCTTCGTCACACAACAGAGAAC
---+-----+-----+-----+-----+-----+-----
TTTCTGCACAGGTAAAGGTCTGACGCTCTCCGAAAAAGAGAAGCAGTGTGTTGTCTCTTG
K D V S I S R L R E A F S L R H T T E N
1300 1320 1340
AAGCCTCACAGCCCAAAGACTCCAGAACCAAGAAGGAGCCCTCTAGGACAGAAAAGGGGT
---+-----+-----+-----+-----+-----+-----
TTCGGAGTGTGCGGTTTCTGAGGTCTTGTTCTTCCTCGGGAGATCCTGTCTTTTCCCCA
K P H S P K T P E P R R S P L G Q K R G
1360 1380 1400
ATGCTGTCTTCTAGCACTTCAGGTGCCATCTCTGACAAAGGCGTCCTGAGACCTCAGAAA
---+-----+-----+-----+-----+-----+-----
TACGACAGAAGATCGTGAAGTCCACGGTAGAGACTGTTCCGCAGGACTCTGGAGTCTTT
M L S S S T S G A I S D K G V L R P Q K
1420 1440 1460
GAGGCAGTGAGTTCCAGTCACGGACCCAGTGACCCTACGGACAGAGCGGAGGTGGAGAAG
---+-----+-----+-----+-----+-----+-----
CTCCGTCACTCAAGGTCAGTGCCTGGGTCACTGGGATGCCTGTCTCGCCTCCACCTCTTC
E A V S S S H G P S D P T D R A E V E K
1480 1500 1520
GACTCGGGGCACGGCAGCACTTCCGTGGATTCTGAGGGGTTTCAGCATCCCAGACACGGGC
---+-----+-----+-----+-----+-----+-----
CTGAGCCCCGTGCCGTCGTGAAGGCACCTAAGACTCCCCAAGTCGTAGGGTCTGTGCCCG
D S G H G S T S V D S E G F S I P D T G
1540 1560 1580
AGTCACTGCAGCAGCGAGTATGCGGCCAGCTCCCCAGGGGACAGGGGCTCGCAGGAACAT
---+-----+-----+-----+-----+-----+-----
TCAGTGACGTCGTCGCTCATACGCCGGTCGAGGGTCCCCTGTCCCCGAGCGTCTTGTA
S H C S S E Y A A S S P G D R G S Q E H
1600 1620 1640
GTGGACTCTCAGGAGAAAGCGCCTGAAACTGACGACTCTTTTTTCAGATGTGGACTGCCAT
---+-----+-----+-----+-----+-----+-----
CACCTGAGAGTCTCTTTTCGCGGACTTTGACTGCTGAGAAAAAGTCTACACCTGACGGTA
V D S Q E K A P E T D D S F S D V D C H

FIG. 3D

19/26

```

1660          1680          1700
.             .             .
TCAAACCAGGAAGATACCGGATGTAAATTTTCGAGTTTTGCCTCAGCCAACTAATCTCGCA
---+-----+-----+-----+-----+-----+-----+-----+-----+
AGTTTGGTCCTTCTATGGCCTACATTTAAAGCTCAAAACGGAGTCGGTTGATTAGAGCGT
S  N  Q  E  D  T  G  C  K  F  R  V  L  P  Q  P  T  N  L  A
1720          1740          1760

.             .             .
ACCCCAAACACAAAGCGTTTTAAAAAGAAGAAATTCTTTCCAGTTCTGACATTTGTCAA
---+-----+-----+-----+-----+-----+-----+-----+-----+
TGGGGTTTGTGTTTCGCAAAATTTTTCTTCTTTAAGAAAGGTCAAGACTGTAAACAGTT
T  P  N  T  K  R  F  K  K  E  E  I  L  S  S  S  D  I  C  Q
1780          1800          1820

.             .             .
AAGTTAGTAAATACTCAGGACATGTCAGCCTCTCAGGTTGATGTAGCTGTGAAAATTAAT
---+-----+-----+-----+-----+-----+-----+-----+-----+
TTCAATCATTTATGAGTCCTGTACAGTCGGAGAGTCCAACTACATCGACACTTTTAATTA
K  L  V  N  T  Q  D  M  S  A  S  Q  V  D  V  A  V  K  I  N
1840          1860          1880

.             .             .
AAGAAAGTTGTGCCCCTGGACTTTTCTATGAGTTCTTTAGCTAAACGAATAAAGCAGTTA
---+-----+-----+-----+-----+-----+-----+-----+-----+
TTCTTTCAACACGGGGACCTGAAAAGATACTCAAGAAATCGATTGCTTATTTTCGTC AAT
K  K  V  V  P  L  D  F  S  M  S  S  L  A  K  R  I  K  Q  L
1900          1920          1940

.             .             .
CATCATGAAGCACAGCAAAGTGAAGGGGAACAGAATTACAGGAAGTTTAGGGCAAAGATT
---+-----+-----+-----+-----+-----+-----+-----+-----+
GTAGTACTTCGTGTCGTTTCACTTCCCCTTGCTCTTAATGTCCTTCAAATCCCGTTTCTAA
H  H  E  A  Q  Q  S  E  G  E  Q  N  Y  R  K  F  R  A  K  I
1960          1980          2000

.             .             .
TGTCCTGGAGAAAATCAAGCAGCCGAAGATGAACTAAGAAAAGAGATAAGTAAAACGATG
---+-----+-----+-----+-----+-----+-----+-----+-----+
ACAGGACCTCTTTTAGTTCGTCGGCTTCTACTTGATTCTTTTCTCTATTCATTTTGCTAC
C  P  G  E  N  Q  A  A  E  D  E  L  R  K  E  I  S  K  T  M
2020          2040          2060

.             .             .
TTTGCAGAAATGGAAATCATTGGTCAGTTTAACCTGGGATTTATAATAACCAAACCTGAAT
---+-----+-----+-----+-----+-----+-----+-----+-----+
AAACGTCTTTACCTTTAGTAACCAGTCAAATTGGACCCTAAATATTATTGGTTTGA CTTA
F  A  E  M  E  I  I  G  Q  F  N  L  G  F  I  I  T  K  L  N

```

FIG. 3E

20/26

2080 2100 2120
GAGGATATCTTCATAGTGGACCAGCATGCCACGGACGAGAAGTATAACTTCGAGATGCTG
---+-----+-----+-----+-----+-----+-----
CTCCTATAGAAGTATCACCTGGTCGTACGGTGCCTGCTCTTCATATTGAAGCTCTACGAC
E D I F I V D Q H A T D E K Y N F E M L
2140 2160 2180
CAGCAGCACACCGTGCTCCAGGGGCAGAGGCTCATAGCACCTCAGACTCTCAACTTAACT
---+-----+-----+-----+-----+-----+-----
GTCGTGCTGTGGCAGAGGTCCCCGTCTCCGAGTATCGTGGAGTCTGAGAGTTGAATTGA
Q Q H T V L Q G Q R L I A P Q T L N L T
2200 2220 2240
GCTGTTAATGAAGCTGTTCTGATAGAAAATCTGGAAATATTTAGAAAGAATGGCTTTGAT
---+-----+-----+-----+-----+-----+-----
CGACAATTACTTCGACAAGACTATCTTTTAGACCTTTATAAATCTTTCTTACCGAAACTA
A V N E A V L I E N L E I F R K N G F D
2260 2280 2300
TTTGTTATCGATGAAAATGCTCCAGTCACTGAAAGGGCTAAACTGATTTCTTGCCAACT
---+-----+-----+-----+-----+-----+-----
AAACAATAGCTACTTTTACGAGGTCAGTGACTTTCCCGATTTGACTAAAGGAACGGTTGA
F V I D E N A P V T E R A K L I S L P T
2320 2340 2360
AGTAAAACTGGACCTTCGGACCCAGGACGTCGATGAACTGATCTTCATGCTGAGCGAC
---+-----+-----+-----+-----+-----+-----
TCATTTTTGACCTGGAAGCCTGGGGTCCTGCAGCTACTTGACTAGAAGTACGACTCGCTG
S K N W T F G P Q D V D E L I F M L S D
2380 2400 2420
AGCCCTGGGGTCATGTGCCGGCCTTCCCGAGTCAAGCAGATGTTTGCCTCCAGAGCCTGC
---+-----+-----+-----+-----+-----+-----
TCGGGACCCCAGTACACGGCCGGAAGGGCTCAGTTCGTCTACAAACGGAGGTCTCGGACG
S P G V M C R P S R V K Q M F A S R A C
2440 2460 2480
CGGAAGTCGGTGATGATTGGGACTGCTCTTAACACAAGCGAGATGAAGAACTGATCACC
---+-----+-----+-----+-----+-----+-----
GCCTTCAGCCACTACTAACCCTGACGAGAATTGTGTTGCTCTACTTCTTTGACTAGTGG
R K S V M I G T A L N T S E M K K L I T

FIG. 3F

21/26

2500 2520 2540
CACATGGGGGAGATGGACCACCCCTGGAAGTGTCCCATGGAAGGCCAACCATGAGACAC
---+-----+-----+-----+-----+-----+-----
GTGTACCCCTCTACCTGGTGGGGACCTTGACAGGGGTACCTTCCGGTTGGTACTCTGTG
H M G E M D H P W N C P H G R P T M R H
2560 2580 2600
ATCGCCAACCTGGGTGTCATTTCTCAGAACTGACCGTAGTCACTGTATGGAATAATTGGT
---+-----+-----+-----+-----+-----+-----
TAGCGGTTGGACCCACAGTAAAGAGTCTTGACTGGCATCAGTGACATACCTTATTAACCA
I A N L G V I S Q N *
2620 2640 2660
TTTATCGCAGATTTTTATGTTTTGAAAGACAGAGTCTTCACTAACCTTTTTTGTTTTAAA
---+-----+-----+-----+-----+-----+-----
AAATAGCGTCTAAAAATACAAAACCTTCTGTCTCAGAAGTGATTGGAAAAACAAAATTT
2680 2700 2720
ATGAAACCTGCTACTTAAAAAAAATACACATCACACCCATTTAAAAGTGATCTTGAGAAC
---+-----+-----+-----+-----+-----+-----
TACTTTGGACGATGAATTTTTTTTATGTGTAGTGTGGGTAAATTTTCTAGAACTCTTG
2740
CTTTTCAAACC
---+-----
GAAAAGTTTGG

FIG. 3G

22/26

YPMS1

m f h h i e n l l i e t e k r c k q e q r y i p v k y l f s m t q l h q i n d i d v h r i t s g q v i t d l t t a v k e l v d n s i d a n a n q i e i i f k d

HMLH2

-----m k q l p a a t v r l l s s s q i i t s v s v k e l i e n s l d a g a t s v d v k l e n

HMLH3

-----i k p i d r k s v h q i c s g q v v l s l s t a v k e l v e n s l d a g a t n i d l k l k d

YPMS1

y g l e s i e c s d n g d g i d p s n y e f l a l k h y t s k i a k f q d v a k v q t l g f r g e a l s s l c g i a k l s v i t t t s p p k - a d k l e y d m v

HMLH2

y g f d k i e v r d n g e g i k a v d a p v m a m k y y t s k i n s h e d l e n l t t y g f r g e a l g s i c c i a e v l t t t r t a a d n n f s t q y v l d g s

HMLH3

y g v d l i e v s d n g c g v e e e n f e g l t l k h h t s k i q e f a d l t q v e t t f g f r g e a l s s l c a l s d v t i s t c h a s a k v g t r l m f d h n

YPMS1

g h i t s k t t t s r n k g t t v l v s q l f h n l p v r q k e f s k t f k r q f t k c l t v i g g y a i i n a a i k f s v w n i t p k g k k n l i l s t m r n

HMLH2

g h i l s q k p s h l g g q t t v t a l r l f k n l p v r k q f y s t a k k c k d e i k i q d l l m s f g i l k p d l r i v f v h n k a v i w q k s r v s d h

HMLH3

g k l i q t p y p r p r g t t v s v q q l f s t l p v r h k e f o r n i k k e y a k m v q v l h a y c i i s a g i r v s c t n q l g q g k r q p v v c t g g s

YPMS1

s s m r k n i s s v f g a g g m r g l e e v d l v l d l n p f k n r m l g k y t d d p d f l d l d y k i r v k g y i s q n s f g c g r n s k d r q f l y v n k r

HMLH2

k m a l m s v l g t a v m n n m e s f q y h s e e s q i y l s g f l p k d a d h s f t s l -----s t p e r s f i f i n s r

HMLH3

p s i k e n i g s v f g q k q l q s l i p f v q l p p s d s v c e e y g l s c s d a l h n l f y i s g f i s q c t h g v g r -----s s t d r q f f f i n r r

FIG. 4A

23/26

YPMS1
HMLH2
HMLH3

PVEYSTLLKCCNEVYKTFNNVQ----FPAVFLNLIELPMSLLIDVNVTPDKRVILLHNERAVIDIFKTTLLSDYNNrqelalp
PVHQDILKLLIRHHYNLCKLKESTRLYPVFFFLKIDVPTADVNVNLTDPDKSQVLLQNKESVLLIALENLMTTCVGGplpstns
PCDPAKVCRLLVNEVYHMYNRHQ----YPFVVVNLNISVDSECVDINVTDPDKRQILLQEEKLLLLAVLKTSLIGMFFDsdvvnkln

YPMS1
HMLH2
HMLH3

krmcsqseqqaqkrlktevfddrstrthesdnenyhtarsesqsnhahfnsttgvidknsngteltsvmdgnytnvtdvig
yennktdvsaadivlsktaetdvlfnkvesgkynsvndtsvipfqndmndesgkntddclnhqisigdfgyghcssei
vsqqplldvegnlikmhaadleekpmvekqdqpslrtgeekdvssrlreaflrhttenkphspktpeprsrplgkkr

YPMS1
HMLH2
HMLH3

secevsdssvvldegnsstptkklpsiktdsqnlslnlnnfsnpefnitspdkaarslekvveepvyfdidggekfqek
snidkntknafqdismsnvswensqteysktcfissvkhtqsgnkdhdidesgeneeeaglenesseisadewsrnlnlk
gmlssstsgaisdkgvlrqpakeavssshgpsdptdraevekdshgtsvdsegfsipdtgshcsseyaaasspgddrgsqe

YPMS1
HMLH2
HMLH3

avlsqadglvfvadnechehtndcchqerrgstdeqdddeadsiyaeiepveinvrtplknsrksiskdnyrslsdglthr
nsvgeniepvkllvpekslpckvsnnyypipeqmnlnedscnkksnvldnksgkvtaaydllsnrvikkpmsasalfvqdh
hvdsgkapetddsfsvdchsnqedtgckfrvlpqpptnlatpntkrfkkeeilssdicqklvntqdmzasasqvavki

FIG. 4B

24/26

YPMS1
HMLH2
HMLH3

kfedell~~eynl~~stknfkeiskngkqms~~si~~iiskrkseageniikndeledfeqqekyltltvskndfkkmevvgqfnlgf
rpqfli~~enp~~kt~~s~~ledatlqieelwktl~~seee~~klkyeekatkdlerynsqmkraieqesqmslkdgrkkikpts~~awn~~laqk
nk~~kv~~vpl~~df~~sm~~s~~lakrikqlhheaqq~~se~~geqnyrkfrakicpgengaedelrkeisktmfaemeiigqfnlgfiitkl

YPMS1
HMLH2
HMLH3

iiivtrk~~v~~dnksdlfi~~vd~~qhas~~de~~kynfetlqavtvfksqk~~li~~ipqpvelsvidelvvdnlpvfekngfklkideeeefg
hklkts~~l~~snqpkldellqsgie~~kr~~rsqnikmvqipfsmkn~~ki~~nfkkqkvdeekdepcli~~hnl~~rfpdawlmstsktevm
nedifi~~v~~dghatdekynfeml~~q~~ghtvlqgqrliapqtl~~nl~~ta~~v~~neavlienleifrkn~~g~~dfvidenapvteraklislp

YPMS1
HMLH2
HMLH3

srvkllslptskqtlfdl~~gd~~fnel~~ih~~likedgglrrdni-----
llnpyrveeallfkrll~~en~~hklpaep~~le~~k~~pi~~mlteslfn~~gsh~~yl~~dv~~lykmtaddqrgsgstylsdprltangfkiklipg
tsknwtfgpqdvdelifmlsds~~pg~~vmc-----

YPMS1
HMLH2
HMLH3

vsitenyleiegmanclpfygvadlkeil~~na~~ilnrna~~ke~~vyecRCSKIRSMFAMRA~~CR~~SSIMIGKPLNKKKT~~M~~TRVVVHNLS

RPSRVKQMFASRA~~CR~~KSVMIGTALNTSEMKKLIT~~TH~~Mg

YPMS1
HMLH2
HMLH3

el~~dk~~pw--NCPHGRPT~~MR~~HLMEL~~ir~~dw~~ss~~fskdyei
hqfgneik~~EC~~VHGRPF~~FH~~HLTYL~~pet~~t-----
emdhpw--NCPHGRPT~~MR~~HIANL~~g~~visqn-----

FIG. 4C

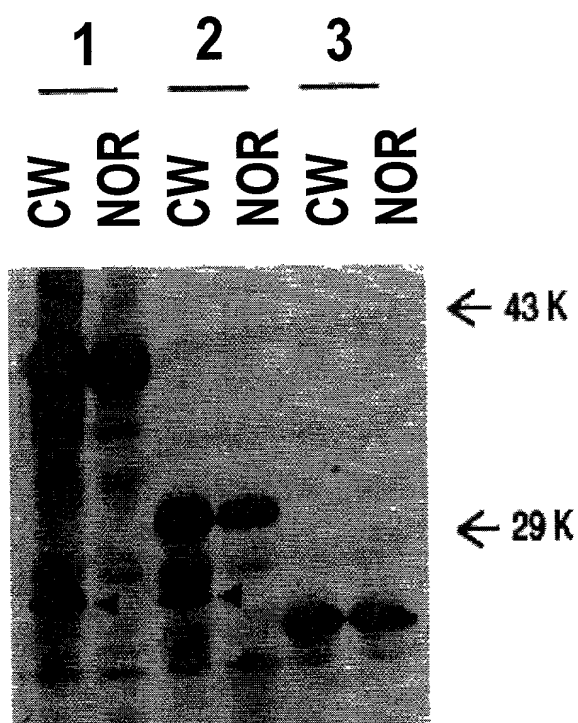


FIG. 5A

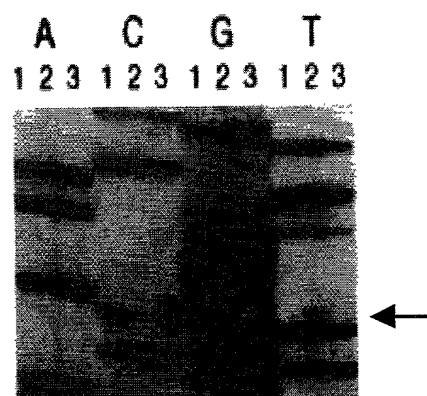


FIG. 5B

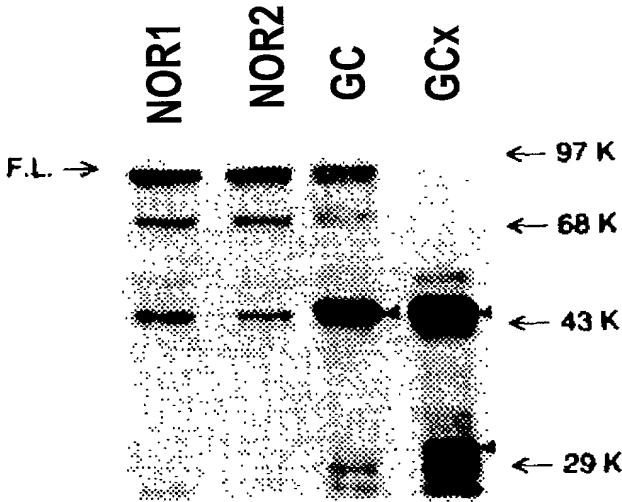


FIG. 6A

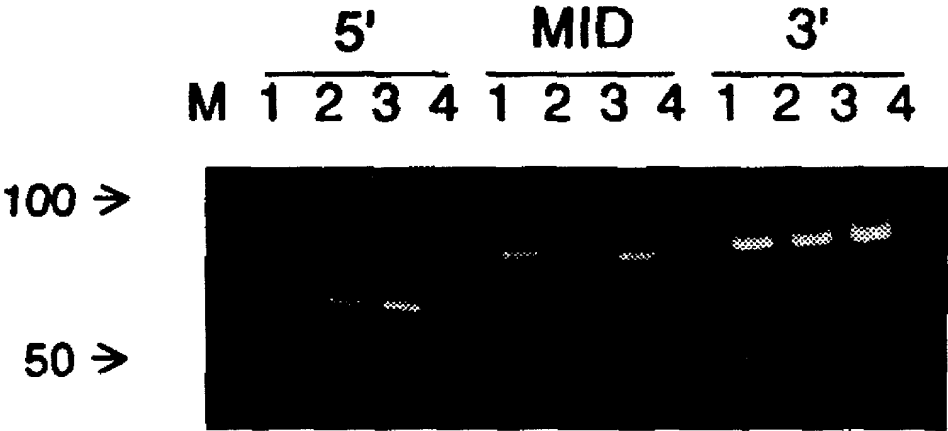


FIG. 6B